

The research program of the Center for Economic Studies (CES) produces a wide range of theoretical and empirical economic analyses that serve to improve the statistical programs of the U.S. Bureau of the Census. Many of these analyses take the form of CES research papers. The papers are intended to make the results of CES research available to economists and other interested parties in order to encourage discussion and obtain suggestions for revision before publication. The papers are unofficial and have not undergone the review accorded official Census Bureau publications. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the U.S. Bureau of the Census. Republication in whole or part must be cleared with the authors.

**OUTPUT PRICE AND MARKUP DISPERSION IN MICRO DATA:
THE ROLES OF PRODUCER AND HETEROGENEITY AND NOISE**

By

Mark J. Roberts*
The Pennsylvania State University
and NBER

and

Dylan Supina* **
University of Vienna

CES 97-10 AUGUST 1997

All papers are screened to ensure that they do not disclose confidential information. Persons who wish to obtain a copy of the paper, submit comments about the paper, or obtain general information about the series should contact Sang V. Nguyen, Editor, Discussion Papers, Center for Economic Studies, Washington Plaza II, Room 211, Bureau of the Census, Washington,

DC 20233-6101, (301-457-1882) or INTERNET address
snguyen@info.census.gov.

Abstract

This paper provides empirical evidence on the extent of producer heterogeneity in the output market by analyzing output price and price-marginal cost markups at the plant level for thirteen homogeneous manufactured goods. It relies on micro data from the U.S. *Census of Manufactures* over the 1963-1987 period. The amount of price heterogeneity varies substantially across products. Over time, plant transition patterns indicate more persistence in the pricing of individual plants than would be generated by purely random movements. High-price and low-price plants remain in the same part of the price distribution with high frequency, suggesting that underlying time-invariant structural factors contribute to the price dispersion. For all but two products, large producers have lower output prices. Marginal cost and the markups are estimated for each plant. The markup remains unchanged or increases with plant size for all but four of the products and declining marginal costs play an important role in generating this pattern. The lower production costs for large producers are, at least partially, passed on to purchasers as lower output prices. Plants with the highest and lowest markups tend to remain so over time, although overall the persistence in markups is less than for output price, suggesting a larger role for idiosyncratic shocks in generating markup variation.

Keywords: Output price; Producer heterogeneity; idiosyncratic shocks

*The authors thank Mike Baye, Keith Crocker, Tim Dunne, Paul Geroski, Zvi Griliches, Adam Jaffe, Robert McGuckin, Pinelopi Goldberg, Dennis Mueller, Peter Reiss, Rafael Rob, John Sutton and the staff of the Center for Economic Studies of the U.S. Census Bureau for helpful comments and discussions on earlier drafts of this paper. This research was conducted at the Center for Economic Studies while the authors were participants in the Census Bureau Research Fellows Program.

* **Dylan Supina would also like to thank the HSR&D Center for Health Care Evaluation, Palo Alto Veterans Affairs Medical Center for partial support.

I. Introduction

Two common features of producer micro data are that, regardless of the dimension examined, producers often differ as much within the same industry as they do across industries and these differences can survive over long periods of time. Besides well-recognized variation in size and capital intensity, recent empirical studies have documented extensive heterogeneity in producer profit rates, failure probabilities, growth rates, the adoption of advanced technologies, R&D expenditures, productivity levels and growth rates, and labor compensation.¹ Coincident with this accumulation of empirical evidence has been the development of theoretical models of firm and industry dynamics capable of explaining some of the empirical patterns. The key components that provide these models with their power to explain the data include: a source of firm-level heterogeneity, often modeled as differences in production efficiency, idiosyncratic shocks that lead to different firm-level learning, investment or

¹ Studies that rely on plant or firm-level panel data sets include: Mueller (1986) on profit differences, Dunne, Roberts, and Samuelson (1989) on growth and failure rates, Davis, Haltiwanger, and Schuh (1996) on growth rates, Lichtenberg and Siegel (1991) and Klette (1996) on R&D investment, Dunne (1994) on technology adoption, Bailey, Hulten, and Campbell (1992), Griliches and Regev (1995), Olley and Pakes (1996), and Tybout (1996) on productivity differences, and Davis and Haltiwanger (1991) on wages.

productivity paths, and sunk entry costs that slow or prevent the sorting out of inefficient from efficient producers.²

While the combination of firm-level heterogeneity in costs, idiosyncratic shocks to firm productivities, and a process of market selection appear to be very important in explaining the dynamic patterns of firm growth and failure, what is missing from the literature is systematic evidence on how these forces affect the output market; specifically, the output prices and markups charged by producers and their evolution over time. This gap in the empirical record can be traced to the difficulty of measuring output prices for individual producers, particularly for representative cross-sections or at more than one point in time. This evidence is also directly relevant to the long-standing question in industrial organization of whether the superior profit performance of some firms reflects more efficient production, the ability to charge higher markups, or a combination of the two (see Demsetz (1973), Peltzman (1977), Mueller (1986), and Schmalensee (1989)). Direct assessment of this issue requires micro data on prices, costs, and markups.

The goal of this paper is to provide empirical evidence on the extent of producer heterogeneity in the output market, specifically, the heterogeneity in output prices and markups. We document both the cross-sectional and time-series patterns in

² Dynamic models of entry and exit with differential efficiency among producers include Jovanovic (1982), Lambson (1991), Hopenhayn (1992), and Ericson and Pakes (1995).

prices and markups at the plant level for thirteen products relying on data drawn from the U.S. Census of Manufactures over the 1963-1987 period. We focus on homogeneous manufactured products in order to remove, as much as possible, price differences due to heterogeneity in the output itself. The data source is unusual in several dimensions. It provides wide coverage of the cross-section of plants in an industry and is not limited to just the large or long-lived producers. It allow us to construct important cost variables, particularly the price of material inputs, at the plant level so that plant-level markups control for cost heterogeneity among producers. Finally, we can trace individual manufacturing plants over time and document the persistence of price, cost, and markup patterns.

A number of earlier studies including Stigler (1961), Stigler and Kindahl (1970), and Pratt, Wise, and Zeckhauser (1979) have documented that the price of a good can vary substantially across different buyers or sellers. Output prices can vary across manufacturing plants in an industry for many reasons including, unobserved product heterogeneity or quality variation across producers, high search costs by consumers, or differences in production costs or competition across local geographic markets. It can also arise in micro data as a result

of reporting or measurement errors.³ Just as the extent of price dispersion is likely to vary across products, no single explanation will be relevant for all the products we study. Instead, the explanation will depend on the particular details of the technology, demand, market competition, and random shocks that pertain to each product. To organize the evidence for the broad range of products we study, we focus on distinguishing price and markup differences at the plant level that persist over time from transitory differences that do not. To the extent that observed price and markup differences persist, it is likely they arise from differences across producers in technology, product quality, demand conditions, market competition, or other structural factors that are constant or change slowly over time. In contrast, if uncorrelated random shocks or measurement errors are the main sources of producer heterogeneity then price and markup differences are much more likely to be transitory.

³ Several recent empirical studies use micro data sets to examine one or more sources of price dispersion. Borenstein and Rose (1994) examine ticket prices on specific airline routes and use differences in demand conditions among the routes to explain differences in the degree of price dispersion. Bresnahan and Reiss (1991) and Dunne and Roberts (1992) use variation in the number of competitors across small geographic markets to explain price differences among tire retailers and bread manufacturers, respectively. Wiggins and Maness (1993) use time-series variation in the number of manufacturers to explain price differences for pharmaceutical products. Shepard (1991) exploits differences in the ability of single and multi-product firms to price discriminate to explain differences in the average price of full and self-serve gasoline among retail service stations in a single geographic market. Abbott (1989) develops a test to distinguish real price dispersion from pure measurement error in the micro data and applies it to output prices for the U.S. cement industry. Van Hoomissen (1988) and Lach and Tsiddon (1992) study the effect of inflation on price dispersion. Pratt, Wise, and Zeckhauser (1979) and Dahlby and West (1986) use models of costly consumer search to explain price dispersion.

The remainder of this paper is organized as follows. The second section discusses some unique characteristics of the *Census of Manufactures* establishment data that are relevant for output price and markup measurement. The third section summarizes the patterns of output price variation across producers, their persistence over time, and the correlation between plant size and prices. The fourth section summarizes the estimation of plant-level marginal cost and the fifth section presents the cross-sectional and time-series patterns in plant markups.

II. Output Price Measurement with Census Establishment Data

In order to make meaningful comparisons of output prices across producers it is important to begin with physically identical goods that are produced by a number of plants. The data are drawn from the U.S. *Census of Manufactures*, which is a survey of all manufacturing establishments conducted in the years 1963, 1967, 1972, 1977, 1982, and 1987. As part of the census, each surveyed establishment reports the value of shipments of each seven-digit SIC product manufactured in the plant as well as the physical quantity produced for a subset of homogeneous products with well-defined units of measure. This allows us to construct a plant-specific average price, as the ratio of the

annual value of shipments to the annual quantity of output shipped, for very disaggregated products in each census year.⁴

In order to compare markups across producers it is necessary to control for cost variation. Since raw materials can often account for more than 50 percent of production costs in a manufacturing plant it is important to control for variation in material prices across establishments. The *Census of Manufactures* also collects data on the plant's expenditure on material inputs and the physical quantity of inputs purchased for detailed raw material categories. This allows us to construct a plant-specific average price of materials for important inputs. The products we have chosen to study all have one or two well-defined material inputs that account for a significant fraction of total material expenditures in the plant and for which expenditure and quantity data are collected.

This study will examine output price dispersion across manufacturing plants for 13 products: white pan bread, roasted coffee, cotton sheeting and allied fabrics, 100% spun polyester blends with cotton, finished wool apparel fabrics, hardwood oak flooring, hardwood plywood, softwood plywood, newsprint,

⁴ This annual average price is a quantity-weighted average of the price of each unit of output sold over the year. Even if all plants sell at exactly the same price at any point in time, we can observe cross-sectional variation in our plant output prices if the common market price fluctuates over the year and plants do not sell identical shares of output at each of the different prices. This type of aggregation bias, however, should not create correlation in a plant's price across census years.

corrugated boxes, motor gasoline, ready-mixed concrete, and tinplate steel cans.⁵ Each of these is a seven-digit SIC product with the exception of coffee, wool fabrics, and corrugated boxes which are aggregates over several seven-digit categories.⁶ Each of the thirteen products is fairly homogeneous and has been chosen to eliminate, as much as possible, price differences that result from output heterogeneity among producers in the seven-digit product category.⁷ With this data we can analyze the cross-sectional distribution of plant-level output prices, shifts of the price distribution over time, and the movement of individual producers through the different yearly distributions.

III. Pricing Patterns at the Plant Level

A. Cross- Section Variation in Output Prices

⁵ The number of products for which quantity data is collected and the number of producers covered have both fallen over time and, since we want to study changes in a producers' price over time, this also affected the set of products we study.

⁶ Coffee and wool fabrics are each an aggregate of two seven-digit products. Corrugated boxes is an aggregate over 9 to 11 seven-digit categories that are defined by the end use of the product.

⁷ Data is available from each of the six censuses between 1963 and 1987 for all products except: cotton sheeting, where 1987 is not available, polyester blends, where 1963 and 1987 are not available, softwood plywood, hardwood oak flooring, and newsprint, where 1982 and 1987 are not available, and tin cans, where data is not available after 1972. The missing years result from redefinition of the seven-digit product codes which made it impossible to construct a consistent definition over all census years. In general, products are aggregated in the later census years. In each case we have constructed consistent product definitions based on the original survey forms from each year and this allows us to avoid problems introduced by large-scale changes in the SIC codes in 1972.

In this section we summarize the heterogeneity in output prices for each product. Table 1 reports three summary statistics of the output price distributions in each year. The first is the trimmed mean, which deletes the lowest five percent and highest five percent of observations in each year in order to remove the effect of outliers in the data. The second measure is the coefficient of variation, the ratio of the standard deviation to the mean multiplied by 100. The final summary statistic is a measure of dispersion that is likely to be more robust to measurement errors and outliers in the data. It is constructed as the difference between the 90th and 10th percentile of the distribution divided by the median. The final column of the table reports the average value of the two dispersion measures across all available years and the products are ranked in the table from highest to lowest dispersion based on the average value of the robust dispersion measure.⁸

The first pattern evident from table 1 is that each of the products is characterized by cross-sectional dispersion, but the magnitude of the dispersion varies substantially across products. At the extremes, the robust dispersion measure averages 2.34 for hardwood plywood and .146 for newsprint. The second pattern is that the amount of dispersion for a given product is similar across the different years. For example, the robust dispersion

⁸ We will maintain this ranking of products from highest to lowest average dispersion in all remaining tables in this paper.

measure for hardwood plywood falls between 2.213 and 2.479 for the different census years, while newsprint varies from .126 to .161. As a result, the ranking of products from high to low dispersion in table 1 would be virtually the same in every year and any shifts in ranking that did occur would be due to small movements in the summary statistic.⁹

The coefficient of variation and robust dispersion measure can be easily compared across industries and across time but they do not provide a complete picture of the price distribution, particularly the higher moments of the distribution or whether there exist different modes. In order to summarize the yearly price distributions more fully, without placing much structure on the data, we construct kernel density estimates for each product in each year.¹⁰ In order to remove the effect of general price-level changes so the separate cross-section price distributions can be compared, each plant's price in year t is normalized by the geometric mean price in year t . Specifically, we estimate the density of $\ln P_{it}^* = \ln(P_{it}) - n^{-1} \ln(J_i(P_{it}))$ where n is the

⁹ The coefficient of variation for softwood plywood is unusually large in one year, 1977, because of a small number of high-price outliers. The average coefficient of variation over 1963-1972 is 17.99 which is in line with the product ranking based on the robust dispersion measure.

¹⁰ Deaton (1989) provides an introduction to kernel density and regression estimators.

number of plants in year t and the term after the minus sign is the log of the geometric mean price over all plants in year t .¹¹

The density of $\ln P^*_{it}$ in year t is estimated as:

$$\hat{d}_h = \sum_i K_h(\ln P^* - \ln P^*_{it})$$

where $\ln P^*$ is an ordered vector of normalized output prices, K_h is the kernel function, and the summation is taken over all plants in year t . For K_h we use the Epanechnikov (1969) kernel:

$$K_h(\ln P^* - \ln P^*_{it}) = \frac{3}{4h} \left(1 - \left(\frac{\ln P^* - \ln P^*_{it}}{h} \right)^2 \right) I(|\ln P^* - \ln P^*_{it}| \leq h)$$

The function $I(\mathcal{C})$ is an indicator function that takes the value one if $\ln P^*$ and $\ln P^*_{it}$ are within an exogenously specified bandwidth h of each other, and zero otherwise. This implies that observations outside of the bandwidth are given a weight of zero while those within the bandwidth are given a weight that varies inversely with their distance from $\ln P^*_{it}$.

Figures 1a-1f graph the estimated densities for $\ln P^*_{it}$ for each year for hardwood plywood, cotton sheeting, bread,

11 This is identical to normalizing by the arithmetic mean of the log prices in each year. An alternative normalization is to subtract the log of the arithmetic mean of the prices. In this data the arithmetic and geometric means of the prices are virtually identical and the choice of normalization has no effect on the results. Our preference for the geometric mean arises because it is less sensitive to outliers than the arithmetic mean.

corrugated boxes, ready-mixed concrete, and gasoline. These products were chosen to span the set from high to low dispersion and because they had sufficient observations to support the nonparametric estimation.¹² The two products with the highest dispersion, hardwood plywood (panel a) and cotton sheeting (panel b), also show the most variation in the densities across years. Even in these cases several similarities are present. The similarity in the overall range of prices across years is clearly evident. Also, a second higher-price mode at approximately 1 (approximately twice the mean price) is evident in each year for hardwood plywood. This suggests that some largely time-invariant characteristic such as product quality may be an important contributor to the high level of price dispersion observed for this product.

The remaining products in panels (c) through (f) of figure 1, show price densities that are more similar across the different years.¹³ In particular, the skewness of the densities for bread and corrugated boxes is present in every year, indicating the consistent presence of some high-price producers.

¹² Four of the products bread, corrugated boxes, concrete, and gasoline have at least 150 observations in each year. The other two products, hardwood plywood and cotton sheeting, have fewer producers and price observations. They each have at least 60 price observations in each year but this may account for the fact that the density estimates are more variable over time for these two products.

¹³ Some of this stability may simply reflect the fact that the sample sizes in every year are larger for these four product than for the products in panels (a) and (b).

Both of these products are sold in small geographic markets and the skewness could arise if there are some local markets with consistently higher prices. Concrete and gasoline are more symmetric distributions and, as seen in table 1, have the lowest overall levels of dispersion among these six products.

Overall, table 1 and figure 1 indicate clearly that, for all of the products, there is cross-sectional variation in plant output prices at every point in time. There are many possible explanations for this including: variation in quality that is not captured by differences in the product definition, differences in production costs among plants that are partly or fully passed on to consumers, high search costs by consumers which can generate price dispersion even with identical costs across producers, differences in the competitiveness of output markets across producers, explicit adjustment costs associated with changing prices that result in producers responding differently to cost or demand shocks, or measurement or reporting errors in the value or quantity of shipments data. As a first step toward understanding the source of the cross-sectional dispersion, we will exploit the multiple time periods in our data and measure to what extent the price differences reflect permanent time-invariant differences among plants versus year-specific idiosyncratic differences.

In order to categorize the likely sources of price variation, assume each plant produces a unique differentiated product, facing a demand elasticity for its product of θ_{it} and marginal cost MC_{it} . The plant chooses price to maximize profit by equating marginal revenue and marginal cost. Both the demand elasticity and marginal cost can vary across plants and time because of market-level time effects (μ_t), plant-specific factors that do not change over time (α_i), and idiosyncratic shocks that are specific to a plant and time period (ε_{it}). The log of the plant's output price can be written as:

$$\ln P_{it} = - \ln (1 - 1/\theta_{it}) + \ln MC_{it} = \mu_t + \alpha_i + \varepsilon_{it} .$$

The market level factors reflect such things as changes in costs due to common input price changes or shifts in market demand for the product. The α_i will reflect factors that vary across plants but are fixed or change very slowly over time. This can include differences in product quality among producers, persistent differences in productive efficiency, or differences in the competitiveness of the local market in which the plant operates. The idiosyncratic shocks will reflect all time-varying factors that alter a plant's marginal cost or demand elasticity as well as measurement or reporting errors in the observed prices. In the case where plants are producing an identical

product and selling it in a single, perfectly competitive output market then observed prices should be explained only by market factors that pertain in that year (μ_t) and measurement or aggregation errors (ϵ_{it}) that result in plant-level price variation. There should be no role for permanent between-plant differences in prices arising from (μ_i) and all cross-sectional variation in prices should reflect measurement and reporting errors. We will assess the importance of between-plant price variation versus within-plant price variation over time with both a variance decomposition and evidence on the persistence of plant-level pricing patterns over time.

The first two columns of table 2 report the amount of between- and within-plant variation in

$\ln P^*_{it}$.¹⁴ Several patterns are evident. First, moving down the table from highest to lowest dispersion products, both variances decline substantially indicating smaller differences in the average price across plants and less time-series variation in prices for each plant. Second, the between-plant variance declines much more rapidly than the within-plant variance as you move from high to low dispersion products. For the three highest dispersion products, hardwood plywood, polyester, and cotton

¹⁴ This variable nets out the geometric mean price in each year so common market-level factors μ_t are not included as a source of price variability. Two products, wool fabrics and tin cans, are not included because of the small number of observations or the short time-series available.

fabric, the between-plant variance is larger than the within-plant variance, indicating that permanent price-level differences across plants are the major source of cross-section price variation. This is consistent with large, permanent differences in demand elasticities or marginal costs across plants. For the remaining products, time-series variation in each plant's price contributes as much (bread, corrugated boxes, and newsprint) or more (coffee, oak flooring, concrete, softwood plywood, and gasoline) to total variation as the between-plant price differences. For these products there is a potentially larger role for time-varying demand or cost conditions at the plant level, as well as noise, to play a role in generating price dispersion.

Table 2 indicates that the cross-sectional dispersion observed in table 1 and figure 1 arises from a different combination of permanent between-plant differences and idiosyncratic factors for each product. These two sources have different implications for the persistence of the plant-level patterns over time. If price heterogeneity arises from structural factors that are slow to change over time, such as the quality of a plant's output or efficiency advantages that the plant may have, then a plant's location in the price distribution should tend to persist over time. In contrast, if the price variability arises solely from plant-level shocks that are not

serially correlated over time, such as idiosyncratic cost or demand shocks or pure measurement errors, then there should be no persistence in the rankings of plants over time. In the next subsection we summarize the time-series persistence as one way of categorizing the importance of structural and random factors in generating price dispersion.

B. Persistence in Pricing Patterns

We explore the persistence in the output price distributions by constructing transition matrices that summarize the movement of plants among different quartiles of the price distribution over time. These transition rates are less likely to be affected by small measurement or reporting errors than the variances reported in Table 2. Table 3 disaggregates plants by the quartile of the price distribution in which they were located in one census year and then reports the proportion that move to each of the other quartiles in the following census year.¹⁵ To simplify the table we have aggregated the second and third quartiles of the price distribution into a single category. The first row of the table indicates that, of all hardwood plywood producers in the highest price quartile in census year t , 83.5 percent remained in the highest price quartile 5 years later, 14.1 percent had dropped to the middle two quartiles of the price distribution and only 2.4 percent had fallen to the lowest quartile. Of plants in the middle two quartiles in year t , 71.9 percent remained there in year $t+1$. Also, 63.2 percent of the low-price plants in year t remained there in $t+1$.

¹⁵ A plant must be in operation in two adjacent census years to be included in these calculations.

As a basis for comparison, if all price variation within a year is due to idiosyncratic shocks, so that $\epsilon_i = 0$, then the probability of being in any quartile in year $t+1$ is independent of the location in year t and the first column of table 3 should be .25, the second column .5, and the final column .25. The same pattern should be observed in columns 4-6 and columns 7-9. We test that each sample proportion is significantly different than .25 or .5, as relevant, and denote cases where the hypothesis is not rejected.

Focusing first on the transition rates of high-price producers in columns 1-3, two patterns are evident. First, the probability of remaining a high-price producer is greater than would be indicated by the random movement of prices. The probability varies from a low of .398 for gasoline to a high of .835 for hardwood plywood. For all of the products the transition rate is significantly different than .25. Second, for many of the thirteen products, the probability of moving into the middle two quartiles does not differ significantly from .5, while the probability of moving into the low price category is substantially less than .25. This implies that the producers that move out of the high-price category predominantly move into the middle of the price distribution but that too few move into the low price quartile for the transitions to be driven solely by

random factors.¹⁶ While plants clearly move through the price distribution over time, the persistence of plants in the high-price part of the distribution suggests that structural, rather than purely idiosyncratic, forces are at work in determining pricing patterns.

The transition rates for plants with prices in the middle two quartiles of the price distribution tell a similar story. The probability of remaining in the middle-price category exceeds .5 for all products, and the difference is significant for eleven of the products, indicating more persistence than would be generated by random price movements. For plants that move out of the middle-price category, the probability of moving to lower prices is greater than the probability of moving to higher prices for 8 of the 13 products. Five of the products have a higher probability of plants moving up in the price distribution and for one product the probabilities are equal. Among the plants in the lowest quartile of the price distribution in year t , between 38.8 percent (gasoline) and 73.3 percent (newsprint) remain in the lowest price quartile 5 years later. All of these transition rates are significantly different than .25. Transitions out of the low-price category are also heavily weighted toward the middle, rather than the upper tail, of the price distribution.

¹⁶ Due to a small number of observations, the probabilities for wool apparel fabrics are not precisely estimated.

Overall, the transition rates indicate more persistence in the ranking of plant prices than is consistent with purely idiosyncratic shocks generating price dispersion and suggest that structural factors that change slowly over time are responsible for some of the price heterogeneity that is observed. In particular, the probability of remaining in the highest or lowest quartile of the price distribution is too large, relative to random chance, and the probability of moving to the other tail of the distribution is too small.

When comparing the probabilities of remaining in the same quartile of the price distribution (cols. 2, 5, and 8) across products, one additional pattern is evident. The degree of persistence is highest for the products with the largest overall levels of price dispersion such as hardwood plywood, the three fabric products, and tin cans. In general, the lower dispersion products such as coffee, oak flooring, concrete, softwood plywood, and gasoline have less persistence and transition rates that more closely approach random noise.¹⁷ This suggests that products with larger price dispersion are not ones with larger idiosyncratic shocks but rather ones where serially-correlated structural factors are more important. The reduction in total

¹⁷ The one prominent exception to this conclusion is newsprint. This product has the lowest overall price dispersion but the highest level of price persistence at the plant level. It suggests that price dispersion arises from small persistent cost, demand, or product quality differences at the plant level and there are very few or very small idiosyncratic shocks.

dispersion as we move down the rows in tables 1 and 2 results from a reduction in the relative importance of permanent to idiosyncratic variation in the price data.

C. Output Prices and Producer Size

Theoretical models of industry evolution often generate differences in producer size with a combination of permanent differences in cost or productivity between producers and idiosyncratic demand or cost shocks over time. The importance of permanent between-plant size differences versus within-plant idiosyncratic variation in size for the products in this study is summarized in the last two columns of Table 2. For every product the between-plant size variation is larger than the total within-plant variation, in some cases by a factor of ten, indicating an important role for permanent, or at least slowly evolving, size differences. If output prices are correlated with plant size or productivity, then the heterogeneity and persistence in output prices documented in the two sections may be another manifestation of the underlying process that generates persistent differences in size.

There are several reasons to suspect that plant size will be systematically related to output price. If larger plants serve larger, more competitive markets then their prices may be lower as a result of the increased competition. If larger plants have lower production costs, either because they are more efficient or can exploit scale economies, these may be passed on to purchasers (Demsetz (1973) and Peltzman (1977)). If larger plants produce very

homogeneous, standardized products while smaller plants manufacture specialty products and serve niche markets then costs and prices may vary with size (Caves and Porter (1977), Porter (1980)).

To summarize the output price-size relationship we regress the log of the plant's normalized output price $\ln P^*_{it}$ on dummy variables for the size quartile in which the plant is located in year t and conduct simple parametric tests of the equality of mean prices across quartiles of the plant size distribution.¹⁸ The regression results and tests for equality of means are reported in table 4. In these regressions $\$_1$ is the intercept and represents the mean price in the smallest size quartile. The other coefficients represent deviations from this for the three larger quartiles. The estimates of $\$_i$ show that, with the exception of gasoline, the smallest producers have prices above the mean. The significant negative estimates for the other coefficients indicate a statistically significant decline in the mean price across the size distribution.¹⁹

¹⁸ The quartiles are constructed from the distribution of real output levels observed across all years. The size distributions are very stable across census years. The regressions are estimated with weighted least squares. The observations within each quartile are used to estimate the standard deviation within the quartile and these are used as quartile-specific weights in the regressions.

¹⁹ The use of quartile regressions is robust to measurement errors in output. Simple regressions of plant price on output are subject to two sources of bias. The first is the standard error-in-variables bias that results from mismeasurement of an explanatory variable and will bias the coefficient on output toward zero. The second bias arises from the fact that price is constructed as the ratio of value of shipments to output so that measurement errors in output are transmitted to price and introduce a spurious negative correlation between price and output. If the true output coefficient is negative then the two biases work in the opposite direction. The magnitude of the biases will decrease as the ratio of the variance of the true output level to the measured output level increases. Because the true cross-sectional size variation is so large in the manufacturing data, and the size differences are so persistent over time, it is unlikely that output measurement errors play a large role in generating the cross-sectional size-

The hypothesis that the mean price is constant across the four quartiles ($\$_2=\$3=\$4=0$) is tested and the F-statistic is reported in column 5 of the table. The hypothesis is rejected for all products except tin cans, gasoline, and newsprint. The latter two are the products with the lowest overall price dispersion.²⁰ The remaining 10 products all show significant price declines with increases in plant size. We next test the hypothesis that the price differences arise only for the smallest plants in the industry by testing that plants in the three largest quartiles have average prices that are equal ($\$2=\$3=\$4$). This hypothesis is rejected for 8 of the 10 products. Finally, the hypothesis that plants in the largest two quartiles have equal prices ($\$3=\4), is rejected for four of the products, hardwood plywood, cotton sheeting, corrugated boxes, and concrete. The results indicate that these four products have systematic price declines that continue across the whole size distribution. Overall in the cross-section, output prices decline with plant size for ten of the thirteen products. Two of the products with no significant size-price correlation, gasoline and newsprint, are the two for which there is very little overall dispersion. These findings further reinforce the conclusion of the previous section that the observed

price relationship which is summarized in table 4. Roberts and Supina (1996) provide pictures of the nonparametric regressions of price on output, which clearly illustrate the negative cross-sectional correlation, for six of the products studied here.

²⁰ Tin cans is unusual because, while there are a large number of plants, they are virtually all owned by a small number of firms. Can plants tend to be constructed near the location of the food processing or beverage plant that will purchase their output. If all cost and price heterogeneity arises at the firm, rather than plant, level and if the size of can plants largely reflects the demand of the near-by users, then there may be persistent price heterogeneity yet no correlation between a plant's size and price.

cross-sectional price variation is not simply random noise. Differences in output prices across plants persist over time and the level of a plant's output price is negatively correlated with the plant's size in the cross-section.²¹ In the next section we examine whether production costs vary systematically across plants as one possible explanation for the cross-sectional pattern of output prices and its persistence over time.

IV. Plant-Level Marginal Cost

An important component of any explanation of price heterogeneity must be heterogeneity in cost, yet cost data is absent from most empirical studies of price dispersion. As part of this research project we have estimated cost function models using the plant-level panel data and constructed estimates of marginal cost that vary by plant and year reflecting differences in factor prices and the level and mix of plant output.²² Several unique characteristics of the census establishment data are important for modeling production costs. First, virtually all plants produce more than one output. Simply ignoring the output of the plant that is not in the product

²¹ This finding also has implications for the measurement of productivity and scale economies with micro data (see Abbott (1992) and Klette and Griliches (1996)). In most applications real output is constructed by deflating the value of firm or plant shipments by a common industry-level price index. Our finding that output prices are lower for large plants implies that the use of a common price deflator will underestimate the output and productivity of large producers relative to small. This problem may be mitigated somewhat in industry-level studies if the output deflator accurately reflects the prices of the large producers since they will account for the bulk of industry output and input use.

²² The complete details of the cost function models are described in Supina (1994) and here we will simply summarize some important aspects of the model and estimation methods.

category we study would lead us to systematically underestimate total plant output and, depending on whether the omitted output was positively or negatively correlated with the output we are interested in, will lead to upward or downward bias in scale parameters.²³ To account for this we aggregate the shipments of all other products manufactured in the plant into a single secondary output, deflate this with an industry-level price index, and include it as an additional control in our multi product cost models. Second, we are able to construct plant-specific prices for materials, labor, and energy and utilize flexible functional forms that allow estimated marginal cost to vary with these factor prices. Third, we have relatively poor information on the plant's capital use. Data for the book value of capital stocks is available by plant in most years but no information on plant service prices or expenditures on capital input is collected. Because of this, cost is defined as the plant's expenditure on materials, labor, and energy inputs.²⁴ Because of the missing capital data we will systematically underestimate the level

²³ In general the omitted and included output will be positively correlated because large plants tend to have higher levels of both outputs. These will lead to an upward bias in the coefficient on output in cost function regressions and thus upward biased estimates of marginal cost.

²⁴ Given the large sample sizes and number of products we examine in this project, it would be an enormous task to construct plant-level service prices for capital that adequately reflected plant-level variation in the opportunity cost of funds, tax variables, and other components of the service price. We will include an ownership-type variable in the cost function model to control for possible differences in capital costs between single and multi-plant producers. Also, because investment flow data is generally only collected for large plants on an annual basis (as part of the *Annual Survey of Manufactures*) it is not possible to construct perpetual-inventory capital stocks for most of the plants in the samples. We do utilize the available information on the book value of the plant's capital in several ways. In the cost model it is used as an instrumental variable for the plant's output and, in the next section, as an additional variable to explain markup differences across plants.

of a plant's marginal cost because the expansion in capital costs with increases in output will not be captured. Perhaps more importantly, since large plants are generally more capital intensive the bias in our estimate of marginal cost will increase with plant size, leading us to potentially underestimate marginal cost and overestimate markups more for large plants.

The basic "long-run" cost model is $C(p_l, p_m, p_e, Q, R, D_t, U)$ where the first three arguments are the prices of labor, material, and energy, Q is the output level of the product we are studying, and R is the output of all other products manufactured in the plant. Cost is defined as the expenditure on labor, materials, and energy.²⁵ Year effects are controlled for with a set of time dummies D_t that distinguish census years. These are interacted with the factor prices to recognize that the factor demand equations may also shift over time. A single-unit/multi-unit ownership dummy U is also included to control for possible differences in capital prices, or differences in the way that central office expenditures are allocated.²⁶

²⁵ If larger plants have more capital intensive technologies, then our inability to measure capital expenditures means that our underestimate of long-run marginal will increase with plant size. This is true assuming capital prices are identical for small and large plants. In practice the price of capital is likely to decline with plant size because large plants are often owned by corporations or large firms who can borrow at lower interest rates. If this is true then the measurement bias could increase or decrease with plant size.

²⁶ We also estimate "short-run" cost models that include the book value of the plant's capital stock as an additional argument, $C(p_l, p_m, p_e, Q, R, D_t, U, K)$. This approach is justified if plants choose labor, materials, and energy inputs to minimize expenditure on these three variable inputs subject to a fixed capital stock. In this case the total cost we measure in the data is the theoretically correct one. Estimates of the short-run model were

The empirical model uses the translog functional form and estimates the cost function together with cost share equations for materials and energy. Estimates of a plant's marginal cost for output Q are constructed as:

$$MC_{it} = \frac{C_{it}}{Q_{it}} \frac{\partial \ln C_{it}}{\partial \ln Q_{it}} = \frac{C_{it}}{Q_{it}} (\beta_Q + \gamma_{QQ} \ln Q_{it} + \gamma_{QR} \ln R_{it} + \sum_j \gamma_{jQ} \ln P_{jit} + \gamma_{QU} U_{it})$$

where the Greek letters are the cost function parameters that are estimated. Estimated marginal cost will vary across plants with differences in factor prices, output levels, and output mix.

Two problems that can arise in production models estimated with micro data are unobserved heterogeneity and measurement error in output. Both can lead to bias in output-related parameters in the cost function and thus bias in estimates of returns to scale or marginal cost (see Griliches (1986), Griliches and Hausman (1986), and Mairesse (1990) for discussion). In general, measurement error in output, which can arise from improper deflation (which is not a problem in our data set) or from deviations between observed and planned output caused by random demand or input supply shocks, will result in a downward bias in output coefficients and an upward bias in estimates of scale economies. Similarly, unobserved efficiency

unsatisfactory because in micro data sets that are dominated by cross-sectional variation among producers, such as the *Census of Manufactures* data, capital acts as a scale parameter like output and reflects long-run differences in plant size, rather than the short-run substitution between fixed and variable inputs. As a result estimates of the cost elasticity with respect to capital were generally positive, reflecting the fact that plants with more capital are larger and have higher variable costs, rather than negative as it would be if it reflected substitution effects.

differences, which will tend to dominate in cross-section data, will generally lead to upward biased estimates of returns to scale, because more efficient producers will be larger and have lower average cost. If we overestimate the degree of scale economies we are likely to underestimate the marginal cost of large producers and thus overestimate their markups.

We estimate the cost model using the seemingly unrelated regression (SUR) estimator. In an attempt to control for measurement error arising from deviations between planned and actual output we use the plant's beginning-of-year capital stock as an instrumental variable for output. Like the SUR estimator, this IV estimator relies primarily on cross-sectional variation in the data. Second, to eliminate plant-specific, time-invariant heterogeneity in the cost function we estimate the cost function in first-differences. Griliches (1986) and Griliches and Hausman (1986) show that, while removing one problem, difference estimators can exaggerate the bias from measurement error. The use of longer time differences, in our case five-year differences, will reduce but not eliminate this problem. To attempt to correct for both sources of bias we also apply the instrumental variable estimator to the time-differenced data using the plant's capital stock as an instrument for the growth in output.

Among the different estimators the SUR and IV estimators that exploit the cross-section variation in the data provide the most

plausible estimates.²⁷ In general the SUR estimates show slight increasing returns to scale across most of the plant size distribution, while the IV estimates indicate constant returns to scale for plants in the largest one or two quartiles of the size distribution. This is consistent with slight measurement error problems in the cross-sectional output data. In the remainder of this paper we will rely on the SUR estimates of the cost function. The main way this will affect our results is to load unobserved heterogeneity effects into the scale-related cost parameters. Given that our interest is in constructing point estimates of marginal cost for each plant that recognize the plants' scale of operation and factor prices, and not separating heterogeneity and scale parameters, this should not be a serious problem.

Table 5 provides summary measures of the variation in estimated marginal cost across the plant size distribution. The format of the table is identical to Table 4. The dependent variable is $\ln MC_{it}^* = \ln(MC_{it}) - n^{-1} \ln(J_i(MC_{it}))$ which is the log of plant-level marginal cost normalized by the log of the geometric mean of marginal cost in the same year. This is

²⁷ The returns to scale estimates from the time-differenced data are implausibly high, which is consistent with differencing exaggerating the measurement error problem in output, and attempts to instrument the growth in plant output are of little value, probably because the instruments are fairly poor. In addition to capital we used lagged values of output, as suggested by Griliches and Hausman (1986), as additional instruments but the results were very similar. As expected, the use of instrumental variables also led to substantial increases in standard errors of the coefficients.

regressed on a set of dummy variables for the quartiles of the size distribution as a way of summarizing the cross-sectional marginal cost variation with plant size.

The marginal cost variation observed in Table 5 is straightforward to summarize for most of the products because it parallels the patterns observed for output price.²⁸ For most of the products, particularly the ones with high price dispersion, marginal cost falls with increases in plant size. Of the first ten products listed in the table, there is a significant decline in marginal cost for eight of them. The same products also had significant declines in output price. Tin cans does not have a significant pattern in marginal cost across the size distribution, but this also parallels the finding of a weak correlation between output price and size in Table 4. Concrete has a slight decline in marginal cost for plants in the middle of the size distribution, but a much more systematic decline in output price. The final two products, gasoline and newsprint, have no systematic output price pattern but both have systematic variation in marginal cost. Overall, for most of the twelve products and for all of the products with the largest price declines, marginal costs are lower for large producers.

This cross-sectional decline in marginal cost reflects a number of factors. Scale effects are important for three of the

²⁸ Wool apparel fabric is not included because the data was too incomplete to estimate the cost model.

products with large cost declines, hardwood plywood, polyester fabric, and cotton sheeting. The scale elasticity, evaluated at the means of the data, equals 1.78, 1.46, and 1.52 for the three products, respectively.²⁹ For a second group of products, bread, oak flooring, concrete, softwood plywood, and newsprint, the scale elasticity at the means of the data varies from 1.13 to 1.20, thus contributing to a decline in marginal cost with plant size. For the final group of products, cans, corrugated boxes, coffee, and gasoline, the scale elasticity lies between 1.02 and 1.04, indicating that increases in plant size have little effect on marginal cost. As discussed above, the scale elasticity will capture both true scale economies as well as the effect of unobserved heterogeneity that results in shifts of the intercept of the cost function across plants. For our purpose, which is to account for across-plant differences in marginal cost due to all sources, this is fine. A second way that scale differences affect our marginal cost estimates is through economies of scope. For virtually all of the products we find significant economies of scope which implies that increases in the plants' other outputs will lower the marginal cost of the product we are focusing on. Since larger plants generally produce more of both outputs Q and R this will contribute to a decline in the

²⁹ The scale elasticity is defined as $SE = (M \ln C / M \ln Q + M \ln C / M \ln R)^{-1}$. Both of the cost elasticities vary with the plant's output mix and factor prices.

marginal cost of output Q as plant size increases. Factor price differences also contribute to heterogeneity in plant marginal costs. Material prices decline with plant size for many of the products, particularly corrugated boxes, coffee, concrete, oak flooring, and softwood plywood, which will lower the marginal cost of larger producers. Countering this is the fact that larger plants almost always pay higher wages. Taken together, these factors contribute to substantial cross-sectional variation in marginal cost, as evidenced by the coefficients in table 5, with the net effect being lower marginal costs for large plants in all but three of the products. This suggests that the lower output prices charged by large plants may reflect underlying cost advantages and this issue is explored in the next section.

V. Plant-Level Markups

In this section we combine the measured output prices and estimated marginal costs into a plant markup and summarize its cross-sectional and time-series patterns. The plant's markup is defined as $\ln(P_{it} / MC_{it})$, which can be interpreted as $-\ln(1 - 1/\theta_{it})$ where θ_{it} is plant i 's demand elasticity in year t . Since both the output price and marginal cost decline with plant size for most products, the cross-sectional variation in the markup will reflect which variable declines more rapidly. Table 6 summarizes the difference in average markups across the

quartiles of the plant size distribution. The format of the table is identical to tables 4 and 5 and the markup is expressed as a deviation from the geometric mean in the relevant year.

There are three reasons we might expect markups to rise with plant size. If larger plants have lower demand elasticities because, for example, there are fewer alternative products available to purchasers, then markups should rise with plant size. If plants have different marginal costs and act as Cournot competitors in a homogeneous-good output market then markups will increase with market share.³⁰ Finally, bias in marginal cost estimates resulting from our inability to measure capital expenditures will generate an increase in the markup with plant size if large plants are more capital intensive.³¹ There is at least one reason to expect markups to decline with plant size. If larger plants operate in more populated geographic markets they are likely to face more competitors than small

³⁰ If all plants are selling an identical product at the same market price, as assumed in the Cournot model, then all the observed cross-sectional variation in output price would have to arise from random measurement errors. This is inconsistent with the persistence patterns documented in table 2 and 3 for most of the products.

³¹ To control for variation in capital use, we also estimate the markup regression including the log or the plant's capital stock and its square as explanatory variables. This controls for the fact that the marginal cost estimates do not capture difference in the capital intensity of the plants and the markup we estimate must, in the long run, cover the plant's capital costs. The capital stock data is not available for 1963 and 1967 and this results in reductions in the sample sizes available for the markup regressions. The effect of including capital in the regressions is discussed below.

geographically-isolated producers and should thus face higher demand elasticities for their product and have lower markups.

The markup-size relationship varies substantially across the 12 products. For six of the products, polyester blend fabrics, bread, coffee, oak flooring, softwood plywood, and newsprint, there is a significant increase in the markup with plant size and the increase is often monotonic across the size class dummies. Two of the 12 products, cotton sheeting and gasoline, have no significant change in markups with plant size. The final four products, hardwood plywood, cans, corrugated boxes and concrete, have markups that decline significantly as plant size increases.

Hardwood plywood is the most unusual case, with a markup that declines by 70 percent as we move from the smallest to largest quartile of plants. This is also the product with the most price dispersion (table 1) and the highest ratio of between-plant to within-plant price variation (table 2). It is likely that, among our set of products, this has the most substantial variation in the quality of output across plants and the price and markup decline is consistent with the large plants specializing in standardized commodity products, which are close substitutes, while smaller plants serve specialized niche markets. The other three products with declining margins, cans, boxes, and concrete, share the common feature that they are sold in relatively small geographic markets. Thus the cross-sectional

markup variation in Table 6 can reflect differences in local markup competition. The decline in both price and markup across the size distribution is consistent with larger plants serving more densely populated, and more competitive geographic markets. We note, however, that the other product sold in small markets, bread, does not show evidence of this competitive effect possibly because there is more product differentiation among bread producers.

The two products with no size related variation in markups, cotton sheeting and gasoline, have the lowest level of markups but very different underlying price and cost patterns. Like hardwood plywood, cotton fabric has a high degree of price dispersion and a high ratio of between-plant to within-plant price variation, suggesting that unobserved quality variation among producers is playing a significant role. The average price of cotton fabric falls by 38 percent from the smallest to largest quartile of plants, but this is closely matched by a decline in marginal cost, again consistent with quality variation, resulting in markups that are small and fairly constant across the range of plant sizes. In the other case, gasoline, there is no systematic price variation and a slight increase in marginal cost, resulting from higher crude oil prices, for larger plants. Markup differences among different size plants appear random.

For the six products with increasing markups, the common thread is that all of them have substantial declines in marginal cost. The coefficient β_4 in Table 5 indicates that, on average, the marginal cost for the largest quartile of producers is between 20 and 66 percent lower than the cost for the smallest quartile of producers. With the exception of newsprint, for which there is little output price variation, these lower costs are partially passed through to the output market in the form of lower prices.³² This pattern is consistent with large producers that face more inelastic demand than their smaller rivals. Their lower costs are reflected partially in lower output prices and partially in higher markups. Two of the products, bread and coffee, are ones where advertising and product differentiation play a role in consumer markets and could lead to more inelastic demand for the output of the large, national producers. The four remaining products are generally sold as intermediate materials to other manufacturing firms and it is not clear without more detailed study of the using industries what to expect about the pattern of demand elasticities faced by large versus small producers.

³² When we also include the plant's capital stock as an additional control variable that may account for rising margins, the markups for these six products rise more substantially than those reported in Table 6. The increased markup for large plants that we observe does not appear to be due to more capital inputs in larger plants.

Differences in demand elasticities and markups across plants can result from idiosyncratic shocks or from more permanent across-plant differences in product quality or other structural factors. We quantify the relative importance of permanent plant-specific factors versus transitory fluctuations in the markup by examining transition rates through the separate markup distributions over time. These are summarized in Table 7, which has the same format as table 3.³³ The first row of the table shows that, of the 25 percent of the plants with the highest markups in year t , .484 of them remained in the highest markup quartile five years later, .387 had moved to the middle two quartiles of the markup distribution, and .129 had moved to the lowest quartile. The table also reports test results for the hypothesis that the transition rates are equal to those that would be generated by purely random movements in markups.

Two patterns are evident. First, as we observed with the price distribution, there is too much persistence in the tails of the markup distribution for the variation to be driven purely by idiosyncratic shocks. For all but one product, softwood plywood, the probability of remaining in the highest markup quartile or remaining in the lowest markup quartile is significantly different than the .25 that would be observed with random movements, and frequently exceeds .50. There is also no obvious

³³ Newsprint is dropped from this table because there were insufficient observations to precisely estimate the transition rates.

difference in the persistence patterns for the products that have lower markups for the large plants (hardwood plywood, cans, corrugated boxes, and concrete). Second, when compared with the transition rates for output prices in table 3, the transitions for plants in the middle of the markup distribution are much more frequently consistent with random movements in markups. That is, we reject the hypothesis of random movements less often with the markups than with the output prices. For five products, polyester blends, cotton sheeting, coffee, oak flooring, and softwood plywood, we cannot reject the null hypothesis of random movements for plants that begin in the middle two quartiles of the markup distribution.

The implication of these patterns is that random shocks or noise in the data are a more important source of markup variation than they are of output price variation but that persistent structural factors are still important in accounting for differences among plants in the high and low markup tails of the distribution. Given our earlier findings on the decline in prices and marginal costs across the size distribution and the importance of permanent between-plant differences in size, the markup transitions are not surprising. If the between-firm variation in prices arises largely from between-firm differences in costs then the markup effectively nets out much of the between-firm variation in these two components and leaves a

larger role for idiosyncratic shocks as the source of markup variation. While the larger idiosyncratic variation makes movements among quartiles of the markup distribution more likely, there is still a permanent component that is identifiable for the highest and lowest markup plants.

VI. Summary and Conclusions

In this paper we exploit data on output prices at the plant-level to study the dispersion of prices, their persistence over time, correlation with producer size, and correlation with production costs for thirteen homogeneous manufactured products. The data reveal that the amount of micro-level price dispersion varies substantially across products but is relatively constant over time for any specific product. Patterns of movement through the price distribution over time reveal that there is more persistence in the pricing of individual plants than would be generated by purely random movements. In particular, high-price plants tend to remain high price and low-price plants remain low price with high frequency, suggesting that underlying time-invariant structural factors contribute to the output price dispersion. For all products except gasoline and newsprint, the two products with the least overall price dispersion in our group, large producers have lower output prices.

To explore if price differentials are cost based we estimate cost function models, construct plant-specific estimates of marginal cost, and examine time-series and cross-sectional patterns of plant-level price-cost markups. The markup remains unchanged or increases with plant size for all but four of the products and declining marginal costs play an important role in generating the markup patterns. The plants with the highest and lowest markups for each product tend to remain in the same location in the cross-sectional distributions in different years, although the degree of persistence for plants in the middle of the distribution is not as strong as it is with the output price, suggesting that idiosyncratic factors are more important in generating markup variation than they are in generating output price variation.

The evidence we offer highlights the heterogeneity in output market performance across different producers. While no manufactured product is perfectly homogeneous across sellers, we have chosen products to come as close as possible to that norm. Even in these cases, there exists substantial and persistent variation in output prices and markups across plants for most of the products we study. This emphasizes that underlying producer heterogeneity in demand elasticities and production costs are

likely to be important determinants of micro level price and markup patterns even in relatively homogeneous goods markets.³⁴

There are, however, differences in the price, cost, and markup relationships across products and we cannot point to a single structural explanation that is consistent with the observations for all. Rather, this diversity suggests that the underlying structural features of the industry differ substantially across products. For example, in corrugated boxes differences in the competitiveness of large and small geographic markets combined with small differences in production costs are consistent with the lower prices and margins for large plants. In contrast, the substantially lower marginal costs of large coffee producers, combined with brand name effects supported through advertising, may be responsible for the slightly lower prices, but significantly higher markups of the large producers. Finally, the small price dispersion, high transition rates, and low correlation of prices, costs, and markups with plant size for the gasoline refiners suggest a process in which random shocks

³⁴ Beginning with Stigler (1961), models with buyer search costs have been used to generate price dispersion for identical products. Reinganum (1979) and Carlson and McAfee (1983) develop models in which marginal cost differences among sellers are reflected in the price distribution. Fishman and Robb (1995) use a combination of search costs, serially-correlated firm cost shocks, and repeat purchases by consumers to generate persistent price and profit differences among producers. At present we do not have any evidence on whether search costs are important or likely to differ across our products and so are not able to draw any conclusions about how relevant these models are for the patterns we observe.

and other sources of noise are the dominant source of output market heterogeneity.

Overall, our empirical findings support the view articulated by Demsetz (1973) that producers differ in their efficiency and that this has implications for how we view the competitiveness of market outcomes. In addition, we provide evidence that these efficiency differences are at least partially reflected in output prices. Large producers not only have lower costs in many product markets, they also have lower output prices. The results suggest the value of further theoretical and empirical work that recognizes the existence of within-industry producer heterogeneity and can link the production and demand sides of the firm's decisions.

References

- Abbott, Thomas A. (1989), *Producer Price Dispersion and the Analysis of Production*, unpublished Ph.D. dissertation, Harvard University.
- Abbott, Thomas A. (1992), "Producer Price Dispersion, Real Output, and the Analysis of Production," *Journal of Productivity Analysis*.
- Baily, Martin, Charles Hulten, and David Campbell (1992), "Productivity Dynamics in Manufacturing Plants," *Brookings Papers on Economic Activity: Microeconomics*, pp. 187-227.
- Borenstein, Severin, and Rose, Nancy L. (1994), "Competitive Price Discrimination in the U.S. Airline Industry," *Journal of Political Economy*, Vol.102, No. 4, pp. 653-683.
- Bresnahan, Timothy F. and Peter C. Reiss (1991), "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, Vol. 99, No.5, pp. 977-1009.
- Carlson, John A. and R. Preston McAfee (1983), "Discrete Equilibrium Price Dispersion," *Journal of Political Economy*, Vol. 91, No. 3, pp. 480-493.
- Caves, Richard E. and Michael E. Porter (1977), "From Entry Barriers to Mobility Barriers," *The Quarterly Journal of Economics*, Vol. 91, No. 2, pp. 241-261.
- Dahlby, Bev and Douglas S. West (1986), "Price Dispersion in an Automobile Insurance Market," *Journal of Political Economy*, Vol. 94, No.2, pp. 418-438.
- Davis, Steven J. and John Haltiwanger (1991), "Wage Dispersion between and within U.S. Manufacturing Plants, 1963-86," *Brookings Papers on Economic Activity: Microeconomics*, Martin N. Bailey and Clifford Winston (eds.), pp. 115-180.
- Davis, Steven J., John Haltiwanger, and Scott Schuh (1996), *Job Creation and Destruction*, Cambridge: MIT Press.
- Deaton, Angus (1989), "Rice Prices and Income Distributions in Thailand: A Non-Parametric Analysis," *Economic Journal*, Vol. 99, pp.1-37.

- Demsetz, Harold (1973): "Industry Structure, Market Rivalry, and Public Policy," *Journal of Law and Economics*, Vol.6 No. 1, pp.1-10.
- Dunne, Timothy (1994), "Plant Age and Technology Use in U.S. Manufacturing Industries," *The Rand Journal of Economics*, Vol. 25, No. 3, pp. 488-499.
- Dunne, Timothy and Mark J. Roberts (1992), "Costs, Demand, and Imperfect Competition as Determinants of Plant-Level Output Prices," in *Empirical Studies in Industrial Organization*, D.B. Audretsch and J.J. Siegfried (eds.), Kluwer Academic Publishers.
- Dunne, Timothy, Mark J. Roberts, and Larry Samuelson (1988), "The Growth and Failure of U.S. Manufacturing Plants," *The Quarterly Journal of Economics*, Vol. 104, No.4, pp. 671-698.
- Ericson, Richard, and Ariel Pakes (1995), "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, Vol. 62, pp. 53-82.
- Fishman, Arthur and Rafael Rob (1995), "The Durability of Information, Market Efficiency and the Size of Firms," *International Economic Review*, Vol. 36, No. 1, pp. 19-36.
- Griliches, Zvi (1986), "Economic Data Issues," *Handbook of Econometrics*, M. Intriligator and Z. Griliches (eds.), Amsterdam: North -Holland.
- Griliches, Zvi, and J.A. Hausman (1986), "Errors in Variables in Panel Data," *Journal of Econometrics*, Vol. 31, pp. 93-118.
- Griliches, Zvi, and Haim Regev (1995), "Firm Productivity in Israeli Industry: 1979-1988," *Journal of Econometrics*, Vol. 65, pp. 175-203..
- Hopenhayn, Hugo (1992), "Entry and Exit in Long Run Equilibria," *Econometrica*, Vol. 60, No. 5, pp. 1127-1150.
- Jovanovic, Boyan (1982), "Selection and the Evolution of Industry," *Econometrica*, Vol. 50, No. 3, pp. 649-670.
- Klette, Tor Jakob (1996), R&D, Scope Economies, and Plant Performance," *The Rand Journal of Economics*, Vol. 27, No. 3, pp. 502-522.

- Klette, Tor Jakob, and Zvi Griliches (1996), "The Inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous," *Journal of Applied Econometrics*, Vol. 11, pp. 343-361.
- Lach, Saul, and Daniel Tsiddon (1992), "The Behavior of Prices and Inflation: An Empirical Analysis of Disaggregated Price Data," *Journal of Political Economy*, Vol. 100, No. 2, pp. 349-389.
- Lambson, Val E. (1991), "Industry Evolution with Sunk Costs and Uncertain Market Conditions," *International Journal of Industrial Organization*, Vol. 9, No. 1, pp.171-196.
- Lichtenberg, Frank and Donald Siegel (1991), "The Impact of R&D Investment on Productivity - New Evidence Using Linked R&D-LRD Data," *Economic Inquiry*, Vol. 29, pp. 203-229.
- Mairesse, Jacques (1990), "Time-Series and Cross-Sectional Estimates on Panel Data: Why are They Different and Why Should They be Equal?," in J. Hartog, G. Ridder, and J. Theeuwes (eds.), *Panel Data and Labor Market Studies*, Elsevier Science Publishers B.V. (North-Holland).
- Mueller, Dennis C. (1986), *Profits in the Long Run*, Cambridge: Cambridge University Press.
- Olley, G. Steven and Ariel Pakes (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, Vol. 64, No. 6, pp. 1263-1297.
- Peltzman, Sam (1977), "The Gains and Losses from Industrial Concentration," *Journal of Law and Economics*, Vol. 20, No. 2, pp.229-263.
- Porter, Michael E. (1980), *Competitive Strategy*, New York: the Free Press.
- Pratt, John W., David A. Wise and Richard Zeckhauser (1979), "Price Differences in Almost Competitive Markets," *Quarterly Journal of Economics*, Vol. 93, pp. 189-211.
- Reinganum, Jennifer F. (1979), "A Simple Model of Equilibrium Price Dispersion," *Journal of Political Economy*, Vol. 87, No. 4, pp. 851-858.

- Roberts, Mark J. and Dylan Supina (1996), "Output Price, Markups, and Producer Size," *European Economic Review Papers and Proceedings*, Vol. 40, Nos. 3/4, pp. 909-921.
- Schmalensee, Richard (1989) "Inter-Industry Studies of Structure and Performance", in *The Handbook of Industrial Organization*, R. Schmalensee and R. Willig (eds.), Amsterdam: North-Holland.
- Shepard, Andrea (1991), "Price Discrimination and Retail Configuration," *Journal of Political Economy*, Vol. 99, No. 1, pp. 30-53.
- Stigler, George (1961), "The Economics of Information," *Journal of Political Economy*, Vol. 64, No. 3, pp. 213-225.
- Stigler, George J. and James K. Kindahl (1970), *The Behavior of Industrial Prices*, New York, Columbia University Press.
- Supina, Dylan (1994), *Price and Markup Dispersion Among U.S. Manufacturing Plants*, unpublished Ph.D. dissertation, The Pennsylvania State University.
- Tybout, James R. (1996), "Heterogeneity and Productivity Growth: Assessing the Evidence," in *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity, and Market Structure*, M. Roberts and J. Tybout (eds.), Oxford: Oxford University Press.
- Van Hoomissen, Theresa (1988), "Price Dispersion and Inflation: Evidence from Israel," *Journal of Political Economy*, Vol. 96, No. 6, pp. 1303-1314.
- Wiggins, Steven N. and Robert Maness (1993), "Price Competition in Pharmaceutical Markets," mimeo.

Table 1**Summary Measures of Output Price Distribution by Year**

Trimmed Mean
(Coefficient of Variation)
[Robust Dispersion Measure]

Product	1963	1967	1972	1977	1982	1987	Mean
Hardwood Plywood (\$/sq. ft. (surface measure))	.255 (80.84) [2.213]	.292 (79.97) [2.270]	.347 (78.89) [2.242]	.548 (90.19) [2.427]	.757 (70.70) [2.479]	.893 (65.60) [2.434]	(77.70) [2.344]
100% Spun Polyester Blend with Cotton (\$/lin. yd.)	--	.555 (58.06) [1.414]	.501 (53.51) [1.041]	.846 (44.01) [1.108]	.779 (59.89) [1.868]	--	(53.87) [1.358]
Cotton Sheeting & Allied Fabrics (\$/lin. yd.)	.300 (52.33) [1.052]	.328 (42.50) [1.091]	.453 (52.50) [1.156]	.804 (78.11) [1.721]	.867 (54.59) [1.728]	--	(56.01) [1.350]
Finished Wool Apparel Fabrics (\$/lin. yd.)	2.249 (34.09) [.931]	2.293 (36.02) [1.017]	2.507 (37.63) [.909]	4.910 (34.03) [.986]	6.721 (18.80) [.671]	7.525 (33.83) [1.191]	(32.40) [.951]
Tinplate Steel Cans (\$1000/base box of steel)	.017 (26.20) [.663]	.019 (30.99) [.667]	.024 (24.96) [.629]	--	--	--	(27.38) [.653]
White Pan Bread (\$/lb.)	.164 (27.36) [.475]	.178 (26.44) [.571]	.198 (25.99) [.493]	.297 (28.02) [.656]	.408 (49.63) [.790]	.475 (33.79) [.850]	(31.87) [.639]
Corrugated Shipping Containers (\$1000/short ton)	.227 (21.79) [.517]	.238 (23.87) [.563]	.256 (36.94) [.613]	.391 (27.73) [.579]	.538 (39.61) [.458]	.621 (37.19) [.505]	(31.19) [.539]
Roasted Coffee (\$/lb.)	.622 (16.41) [.409]	.702 (15.01) [.392]	.846 (14.35) [.365]	2.955 (16.08) [.418]	2.290 (17.49) [.394]	2.601 (25.121) [.539]	(17.41) [.420]
Hardwood Oak Flooring (\$/board ft.)	.147 (13.84) [.312]	.155 (34.48) [.394]	.204 (15.86) [.282]	.419 (85.67) [.685]	--	--	(34.50) [.412]
Ready-Mixed Concrete (\$/cubic yd.)	.014 (20.30) [.328]	.015 (13.22) [.318]	.018 (15.25) [.370]	.028 (13.59) [.321]	.043 (17.38) [.355]	.046 (37.20) [.425]	(19.49) [.353]
Interior Softwood Plywood (\$/sq. ft. 3/8" basis)	.058 (22.20) [.316]	.057 (14.66) [.371]	.088 (17.11) [.389]	.152 (75.33) [.202]	--	--	(32.33) [.320]
Motor Gasoline (\$/barrel)	4.830 (11.50) [.307]	5.108 (11.62) [.304]	5.337 (11.75) [.261]	15.606 (6.36) [.159]	39.591 (6.22) [.133]	22.873 (8.17) [.167]	(9.27) [.222]
Newsprint (\$1000/short ton)	.122 (7.41) [.126]	.126 (8.16) [.157]	.147 (5.34) [.161]	.280 (4.48) [.140]	--	--	(6.35) [.146]

Table 2

Between and Within-Plant Variation in Price and Output

(Output price expressed as deviation from year mean)

Product	Between-Plant Variance ln P*	Within-Plant Variance ln P*	Between-Plant Variance ln Q	Within-Plant Variance ln Q
Hardwood Plywood	.294	.149	1.155	.325
100% Spun Polyester Blends with Cotton	.141	.125	1.493	1.228
Cotton Sheeting & Allied Fabrics	.162	.119	1.422	.889
White Pan Bread	.044	.045	2.510	.343
Corrugated Shipping Containers	.031	.032	1.389	.287
Roasted Coffee	.017	.026	2.440	.220
Hardwood Oak Flooring	.020	.049	.871	.730
Ready-Mixed Concrete	.015	.020	.856	.280
Softwood Plywood	.012	.021	1.366	.768
Motor Gasoline	.004	.006	2.745	.261
Newsprint	.002	.002	.887	.344

Table 3

Transition Rates Across Output Price Quartiles

Product	Plants in the highest price quartile in year t			Plants in the middle two price quartiles in year t			Plants in the lowest price quartile in year t		
	Highest quartile in $t+1$	Middle two quartiles in $t+1$	Lowest quartile in $t+1$	Highest quartile in $t+1$	Middle two quartiles in $t+1$	Lowest quartile in $t+1$	Highest quartile in $t+1$	Middle two quartiles in $t+1$	Lowest quartile in $t+1$
Hardwood plywood	.835	.141	.024	.110	.719	.171	.039	.329	.632
100% Spun polyester blends with cotton	.426	.519 ^a	.055	.164	.746	.090 ^b	.135	.308	.558
Cotton sheeting and allied fabrics	.538	.404 ^a	.058	.173	.655	.173	.063	.266	.672
Finished wool apparel fabrics	.529	.352 ^a	.118 ^b	.206 ^b	.647 ^a	.147 ^b	.052	.368 ^a	.579
Tinplate steel cans	.610	.293	.098	.176	.631	.193	.046	.402 ^a	.552
White pan bread	.501	.374	.125	.148	.637	.216	.083	.382	.535
Corrugated shipping containers	.501	.423	.075	.187	.619	.194	.073	.396	.531
Roasted coffee	.429	.438 ^a	.133	.174	.601	.225 ^b	.138	.459 ^a	.404
Hardwood oak flooring	.521	.375 ^a	.104	.182 ^b	.625	.193 ^b	.130	.457 ^a	.413
Ready-mixed concrete	.505	.407	.088	.196	.593	.212	.099	.426	.475
Interior softwood plywood	.483	.449 ^a	.067	.181	.559 ^a	.260 ^b	.110	.425 ^a	.466
Motor gasoline	.398	.393	.209 ^b	.220 ^b	.573	.207	.146	.466 ^a	.388
Newsprint	.538	.385 ^a	.077	.193 ^b	.710	.097	.000	.267	.733

a - do not reject that the proportion equals .5 at the .05 significance level, b - do not reject that the proportion equals .25 at the .05 significance level

Table 4
Output Price Differences by Size Quartile
(standard errors in parentheses)
Dependent Variable: $\ln P_{it}^*$

<i>Product</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Obs.</i>	<i>R</i> ²
Hardwood plywood	.522** (.055)	-.415** (.078)	-.733** (.074)	-.936** (.066)	60.95** (2.0)	31.49** (.9)	10.88** (.6)	491	.33
100% Spun polyester blends with cotton	.200** (.056)	-.083 (.072)	-.258** (.067)	-.459** (.069)	15.43**	19.20**	13.31**	367	.15
Cotton sheeting and allied fabrics	.213** (.056)	-.150** (.071)	-.324** (.073)	-.379** (.076)	8.26**	6.88**	.65	337	.09
Finished wool apparel fabrics	.281** (.069)	-.338** (.090)	-.343** (.082)	-.436** (.085)	7.40**	1.26	2.04	94	.25
Tinplate steel cans	.037 (.025)	-.037 (.032)	-.061** (.029)	-.049* (.030)	1.32	.40	.24	712	.01
White pan bread	.091** (.011)	-.088** (.014)	-.136** (.013)	-.139** (.013)	35.21**	11.90**	.06	2807	.05
Corrugated shipping containers	.136** (.009)	-.119** (.011)	-.185** (.010)	-.240** (.010)	242.09**	138.51**	83.28**	4571	.18
Roasted coffee	.047** (.013)	-.021 (.018)	-.079** (.019)	-.089** (.019)	8.28**	8.51**	.31	762	.04
Hardwood oak flooring	.107** (.040)	-.122** (.044)	-.131** (.046)	-.175** (.048)	3.98**	1.57	1.74	280	.05
Ready-mixed concrete	.046** (.003)	-.028** (.004)	-.059** (.004)	-.096** (.005)	122.05**	119.92**	71.50**	11360	.04
Interior softwood plywood	.059** (.029)	-.032 (.031)	-.091** (.031)	-.113** (.030)	15.69**	17.17**	2.59	418	.13
Motor gasoline	-.004 (.006)	.003 (.009)	.008 (.009)	.005 (.009)	.23	.21	.16	964	.001
Newsprint	.024 (.024)	-.038 (.026)	-.034 (.027)	-.025 (.026)	.81	.39	.26	67	.05

** Significant at the .05 level, * Significant at the .1 level

Table 5
Marginal Cost Differences by Size Quartile
(standard errors in parentheses)
Dependent Variable: $\ln MC_{it}^*$

<i>Product</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Obs.</i>	<i>R</i> ²
Hardwood plywood	.201** (.064)	-.326** (.092)	-.259** (.106)	-.214** (.081)	3.34** (.420)	0.89** (.39)	0.34 (.34)	272	.05
100% Spun polyester blends with cotton	.387** (.066)	-.354 (.082)	-.525** (.075)	-.665** (.076)	26.54**	13.12**	7.61**	268	.29
Cotton sheeting and allied fabrics	.274** (.065)	-.237** (.077)	-.413** (.080)	-.448** (.086)	9.15**	6.04**	0.23	297	.11
Tinplate steel cans	-.034 (.035)	.022 (.044)	.035 (.042)	.081** (.041)	1.56	1.92	2.19	669	.01
White pan bread	.135** (.011)	-.131** (.015)	-.182** (.015)	-.231** (.016)	61.59**	25.13**	11.85**	2638	.09
Corrugated shipping containers	.026** (.012)	-.028** (.014)	-.026** (.013)	-.052** (.013)	6.33**	6.11**	10.56**	4393	.01
Roasted coffee	.113** (.022)	-.101** (.028)	-.127** (.028)	-.222** (.030)	13.61**	10.68**	12.48**	637	.08
Hardwood oak flooring	.231** (.063)	-.195** (.078)	-.352** (.086)	-.374** (.084)	6.22**	3.73**	0.07	200	.11
Ready-mixed concrete	.009 (.010)	-.030** (.012)	-.022* (.012)	.015 (.012)	6.45**	12.29**	16.02**	6045	.01
Interior softwood plywood	.384** (.057)	-.357 (.062)	-.511** (.060)	-.662** (.063)	48.92**	33.72**	20.33**	167	.55
Motor gasoline	-.053** (.016)	.050** (.018)	.075** (.018)	.084** (.018)	7.74**	3.79**	0.54	918	.03
Newsprint	.098** (.044)	-.041 (.066)	-.158 (.085)	-.201** (.060)	3.26**	3.09*	0.40	47	.24

** Significant at the .05 level, * Significant at the .1 level

Table 6
Markup Differences by Size Quartile
(standard errors in parentheses)
Dependent Variable: $\ln(P_{it}/MC_{it})^*$

<i>Product</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Obs.</i>	<i>R</i> ²
Hardwood plywood	.392** (.054)	-.306** (.073)	-.551** (.088)	-.700** (.064)	34.07** (2.20)	20.99** (2.29)	3.38* (.38)	272	.34
100% Spun polyester blends with cotton	-1.62** (.070)	.212** (.078)	.197** (.076)	.237** (.076)	3.72**	0.43	0.85	268	.05
Cotton sheeting and allied fabrics	.014 (.070)	.029 (.079)	-.027 (.081)	-.056 (.077)	0.82	1.55	0.32	297	.01
Tinplate steel cans	.041 (.028)	-.032 (.034)	-.054 (.034)	-.078** (.032)	2.24*	1.76	1.01	669	.01
White pan bread	-.057** (.012)	.054** (.016)	.069** (.016)	.108** (.018)	10.69**	6.84**	6.42**	2638	.02
Corrugated shipping containers	.112** (.009)	-.096** (.012)	-.164** (.010)	-.189** (.010)	149.85**	66.99**	15.55**	4393	.12
Roasted coffee	-.050** (.023)	.045 (.029)	.041 (.029)	.115** (.031)	3.73**	4.56*	7.65**	637	.02
Hardwood oak flooring	-.150** (.067)	.102 (.080)	.257** (.084)	.239** (.084)	3.37**	3.23*	0.07	200	.06
Ready-mixed concrete	.027** (.010)	.013 (.012)	-.022* (.011)	-.099** (.011)	44.66**	78.23**	77.83**	6045	.03
Interior softwood plywood	-.303** (.067)	.289** (.071)	.397** (.071)	.522** (.072)	29.63**	25.93**	14.42**	167	.42
Motor gasoline	.019 (.015)	-.023 (.019)	-.029 (.018)	-.022 (.019)	0.72	0.15	0.24	918	.00
Newsprint	-.086** (.042)	.016 (.059)	.140* (.076)	.187** (.058)	3.53**	4.52*	.39	46	.25

** Significant at the α .05 level, * Significant at the α .1 level

Table 7

Transition Rates Across Quartiles of the Markup Distribution

Product	Plants in the highest markup quartile in year t			Plants in the middle two markup quartiles in year t			Plants in the lowest markup quartile in year t		
	Highest quartile in $t+1$	Middle two quartiles in $t+1$	Lowest quartile in $t+1$	Highest quartile in $t+1$	Middle two quartiles in $t+1$	Lowest quartile in $t+1$	Highest quartile in $t+1$	Middle two quartiles in $t+1$	Lowest quartile in $t+1$
Hardwood plywood	.484	.387 ^a	.129	.111	.778	.111	.071	.214	.714
100% Spun polyester blends with cotton	.469	.438 ^a	.094	.270 ^b	.556 ^a	.175 ^b	.154 ^b	.308	.538
Cotton sheeting and allied fabrics	.467	.400 ^a	.133 ^b	.174 ^b	.581 ^a	.244 ^b	.022	.457 ^a	.522
Tinplate steel cans	.563	.350	.088	.152	.591	.258 ^b	.083	.524 ^a	.393
White pan bread	.541	.363	.096	.164	.631	.204	.098	.376	.526
Corrugated shipping containers	.491	.410	.099	.133	.611	.255 ^b	.075	.487 ^a	.438
Roasted coffee	.505	.398	.097	.250 ^b	.543 ^a	.207 ^b	.094	.406 ^a	.500
Hardwood oak flooring	.476	.429 ^a	.095	.213 ^b	.596 ^a	.191 ^b	.037	.333 ^a	.630
Ready-mixed concrete	.396	.476 ^a	.129	.184	.614	.202	.096	.362	.542
Interior softwood plywood	.471 ^b	.353 ^a	.176 ^b	.125	.563 ^a	.313 ^b	.154 ^b	.461 ^a	.385 ^b
Motor gasoline	.456	.443 ^a	.101	.205	.573	.223 ^b	.110	.433 ^a	.457

a - do not reject that the proportion equals .5 at the .05 significance level, b - do not reject that the proportion equals .25 at the .05 significance level